Published in final edited form as:
Hear Res. 2013 April ; 298: 36–48. doi:10.1016/j.heares.2013.01.011.

Effect of acoustic similarity on short-term auditory memory in the monkey

Brian H. Scott[a], **Mortimer Mishkin**[a], and **Pingbo Yin**[a,b]

Brian H. Scott: brianscott@mail.nih.gov; Mortimer Mishkin: mishkinm@mail.nih.gov; Pingbo Yin: pyin@umd.edu

[a]Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, 49 Convent Drive, Room 1B80, Bethesda, MD 20892

[b]Neural Systems Laboratory, Institute for Systems Research, University of Maryland, College Park, MD 20742

Abstract

Recent evidence suggests that the monkey's short-term memory in audition depends on a passively retained sensory trace as opposed to a trace reactivated from long-term memory for use in working memory. Reliance on a passive sensory trace could render memory particularly susceptible to confusion between sounds that are similar in some acoustic dimension. If so, then in delayed matching-to-sample, the monkey's performance should be predicted by the similarity in the salient acoustic dimension between the sample and subsequent test stimulus, even at very short delays. To test this prediction and isolate the acoustic features relevant to short-term memory, we examined the pattern of errors made by two rhesus monkeys performing a serial, auditory delayed match-to-sample task with interstimulus intervals of 1 s. The analysis revealed that false-alarm errors did indeed result from similarity-based confusion between the sample and the subsequent nonmatch stimuli. Manipulation of the stimuli showed that removal of spectral cues was more disruptive to matching behavior than removal of temporal cues. In addition, the effect of acoustic similarity on false-alarm response was stronger at the first nonmatch stimulus than at the second one. This pattern of errors would be expected if the first nonmatch stimulus overwrote the sample's trace, and suggests that the passively retained trace is not only vulnerable to similarity-based confusion but is also highly susceptible to overwriting.

Keywords

rhesus; macaque; primate; working memory; sound; vocalizations

1. Introduction

Studies of auditory memory in nonhuman primates consistently report extremely slow learning of the rule for delayed match-to-sample (D'Amato et al., 1985; Fritz et al., 2005; Wright, 1999) and short sample-stimulus forgetting thresholds (~ 30 s), whether the task utilizes only two sounds (Colombo et al., 1996) or trial-unique sounds (Fritz et al., 2005). These findings suggest that although monkeys are easily able to form long-term memories in vision and touch (Mishkin, 1978; Murray et al., 1983), they may be unable to do so in audition, and are therefore limited acoustically to short-term memory (Fritz et al., 2005). More recently, we obtained evidence that even this type of auditory memory in the monkey

Corresponding Author: Brian Scott, 49 Convent Drive, Room 1B80, Bethesda, MD 20892 USA, brianscott@mail.nih.gov, Phone: (301) 443-7867, Fax: (301) 402-0046.

Supporting Information: none

is sharply limited (Scott et al., 2012), as it is likely to be dependent on a passive form of short-term memory (pSTM). This passive form can be distinguished from the active form (viz., working memory, WM) in that it relies exclusively on passively retained sensory traces rather than on activation of previously stored neural representations either of particular sounds or of sound categories, e.g. tones, vocalizations, environmental sounds, etc.

The proposition that monkeys may lack auditory long-term memory (LTM), and by extension WM, may appear to be inconsistent with the monkey's ability to react appropriately to species-specific communication calls, or to learn auditory discrimination tasks by instrumental conditioning. However, the first of these behavioral abilities is likely to rely instead on cross-modal association (in which a call activates the stored representation of a visual associate), and the second, on the formation and strengthening of stimulus-response habits, with neither of them depending on auditory LTM *per se* (Scott et al. 2012). Our definition of auditory LTM requires that a current sound be recognized, i.e. that it reactivate the stored representation of the same sound heard previously, as demonstrated by delayed matching-to-sample.

In an earlier study, we tested auditory STM in two rhesus monkeys using a serial delayed-match-to-sample (DMS) task (Scott et al., 2012). Two lines of evidence supported the proposal that the monkeys' performance relied on a pSTM trace rather than on a more robust representation retrieved from long-term memory. First, performance was particularly poor for a match stimulus that followed the nonmatch 'distracters', indicating that the memory trace was fragile and so was easily overwritten by subsequent stimuli (i.e., highly susceptible to retroactive interference). Second, this low level of performance prevailed despite a task design in which the nonmatch stimuli were drawn from sound categories different from that of the sample, so that simply matching to category would have enabled perfect performance.

In fact, the monkeys' DMS performance did show an effect of sound category, but in a counter-intuitive direction: Performance was better for tones and narrow band-passed noise stimuli than for natural sounds, including vocalizations. Thus, under our task conditions, ethological significance of the stimuli did not seem to be a relevant factor in the monkeys' performance, leading us to speculate that their delayed matching was based solely on the sensory qualities of the stimuli. If so, then degree of acoustic similarity between sample and test items should predict DMS performance, and focusing the analysis on this variable should lead to identification of the relevant acoustic feature(s) for which sample-test similarity predicts the behavioral outcome.

The present study addressed this hypothesis by examining the patterns of errors made by our subjects over many tens of thousands of trials of auditory DMS. The analysis revealed that their errors resulted primarily from confusion between pairs of sounds with similar spectral content independent of the degree of their temporal-envelope similarity. These findings suggest that the monkey's short-term memory is based solely on passive retention of an acoustic trace dominated by spectral content, and this impoverished trace could conceivably reflect a limitation of auditory memory among nonhuman primates generally.

2. Methods

2.1. Subjects and Apparatus

Subjects were two adult male rhesus monkeys (*Macaca mulatta*). One monkey (F) was naïve prior to this study, whereas the other monkey (S) had been trained in an earlier study on an auditory discrimination task (Yin et al., 2008); the possible influence of that training on monkey S's performance in the present study is discussed below (Discussion, 4.2). Testing took place within a double-walled, sound-attenuating booth (IAC, Bronx NY), with the monkey seated in a primate chair fitted with a metal contact bar. A sipper tube was positioned for delivery of liquid reward (typically water) under computer control (Crist Instruments, Hagerstown, MD). Because the behavioral task was coupled intermittently with electrophysiological recording sessions, the monkey's head position was fixed during testing by a titanium head-holder secured to the primate chair. The data in this report, however, were collected during daily sessions when only behavioral testing was conducted.

The behavioral task was controlled by NIMH Cortex software (Laboratory of Neuropsychology, NIMH; http://dally.nimh.nih.gov/), which triggered sound playback via a custom-built interface with a second computer running SIGNAL software (Engineering Design, http://www.engdes.com/). The output of the SIGNAL buffers was flattened across frequency (Rane RPM 26v parametric equalizer, Mukilteo WA), attenuated (Agilent HP 355C and 355D), amplified (NAD, Pickering, Ontario), and delivered via a loudspeaker (Ohm Acoustics, NY) located 1 m directly in front of the animal's head. Sound level was calibrated with a Brüel and Kjær 2237 sound-level meter using A-weighting. Task-relevant events were collected on a CED 1401 acquisition system controlled by Spike2 software (Cambridge Electronic Design, UK). Data were exported to MATLAB (Mathworks, Natick, MA) for analysis, and statistics were computed by the MATLAB Statistics Toolbox.

2.2. Delayed match-to-sample task

Preliminary training on the DMS rule was described in the earlier study (Scott et al., 2012). Once the rule was acquired, the task proceeded as follows. The animal initiated a trial by holding a contact bar for 300 ms (Fig. 1A). This triggered presentation of a sample stimulus (~300 ms in duration and drawn randomly from a set of 21 stimuli; see below), followed by 1–3 test sounds with a variable interstimulus interval (ISI) of 800–1200 ms. When the test sound was the same as the sample (a match), the animal was required to release the bar within a 1200-ms response window beginning 100 ms after the onset of the match sound. A correct response (a "hit") earned a few drops (0.3–0.5 mL) of liquid reward after bar release. A response within the first 100 ms following match onset was considered an "early-release" error. Failure to release the bar by the end of the response window was counted as a "miss" error. If the test sound was a nonmatch, the animal was required to hold the bar (a "correct rejection") until the match stimulus was presented. Release to the nonmatch stimulus was counted as a "false alarm" (FA) error. Any type of error aborted the trial and was penalized by a 3-s timeout in addition to the standard 3-s intertrial interval; the penalty was intended to discourage animals from aborting trials with multiple nonmatches. Each trial ended after release of the bar, but if the bar was released during stimulus presentation, the full stimulus played out before the trial was reset. Trials with zero, one, or two nonmatch sounds were randomly generated with equal probability. In an attempt to reduce the memory demands of the DMS task, the nonmatch stimuli were always drawn from categories different from that of the sample, which were otherwise selected randomly on each trial. Trials were organized in blocks such that each stimulus in the set served as the sample in a pseudorandom order before the same stimulus appeared as the sample again.

2.3. Stimuli

The set of 21 sounds is illustrated in Fig. 1B. All sounds were recorded at 16-bit resolution at a sampling rate of 32 kHz, except for the Mvocs, for which the sampling rate was 24 kHz. The rhesus vocalizations were collected from a colony on Cayo Santiago, Puerto Rico (provided courtesy of Marc Hauser), so the individual callers were unfamiliar to our two subjects. All stimuli were equalized in root-mean-square amplitude to have approximately equal loudness and were presented at 60–70 dB SPL.

In a control experiment, designed to determine which stimulus dimension (spectral or temporal) was the more important for performance, we used a version of the stimulus set in which the sounds were manipulated to contain information in only one or the other dimension. These data were collected in a separate block of sessions after collection of the DMS data described above. The ‘temporal-only’ stimuli were constructed by applying the envelope of the original sounds (as extracted by the Hilbert transform) to Gaussian noise. The ‘spectral-only’ stimuli were generated by measuring the frequency spectrum of the original sounds (power spectral density by the Welch method, 50% overlap, 64 sample segment length, Hamming window) and constructing a noise stimulus with the same spectrum. At each frequency ≥ 60 Hz, a sine function of random phase was generated with an amplitude proportional to the power spectral density at that frequency; the summed signal had a flat envelope (300-ms duration, with a 10-ms linear on/off ramp) and was normalized in root-mean-square amplitude to the original sound. The spectra of the resulting stimulus and the original stimulus were overlaid to confirm that they were spectrally identical. Some sounds in the original set had identical temporal envelopes or spectra, so the redundant stimuli were removed to leave only unique sounds (N = 13 temporal-only and 18 spectral-only).

2.4. Behavioral analysis

As described in detail elsewhere (Scott et al., 2012; Yin et al., 2010), performance on the DMS task was measured both by percent correct and by the Discrimination Index (DI) derived from signal detection theory, which incorporates both the accuracy and reaction time (RT) of the behavioral response. The bar-release latency within the response window was measured relative to the onset of the test sound that elicited the release, which was scored as a hit if that sound was the match, or a FA if that sound was a nonmatch. (Release to the sample was considered an aborted trial and discarded.) The cumulative probabilities of hits and FAs were then calculated at 50-ms intervals across the response window. The cumulative hit and FA probabilities, plotted against one another, define a curve in ROC space, and DI is measured as the area under the curve (ROC value). Perfect performance would yield a DI value of 1, whereas a random response would yield a value around 0.5. To derive a threshold for above-chance performance, the matrix of hit and false-alarm labels was randomly shuffled with respect to the corresponding RTs, and the DI was computed from the shuffled data. This computation was repeated 100 times, and the threshold was defined as 2 standard deviations (SDs) above the mean of the shuffled DIs.

Figure 2A presents a schematic diagram of the three trial types (i.e., trials with either zero, one, or two nonmatch stimuli), which were randomly interleaved in the task, and also shows the positions in the sequence at which sample (S), match (M), and nonmatch (NM) stimuli could appear. The sound at position 1 was always the sample; the sound at position 2 could be a match or nonmatch; if it was a nonmatch (and the animal successfully withheld response), another sound was presented at position 3, and this could also be a match or nonmatch; finally, if the sound at position 3 was a nonmatch (and the animal again withheld response), the sound presented at position 4 would always be a match. The DI measure includes hits and FAs from stimulus positions 2 and 3, and these are combined unless stated otherwise; position 4 was excluded from the DI measure, because the stimulus at this position was always a match, and therefore no FA was possible.

Performance was also assessed by the FA rate, calculated as FA/(FA+CR), the ratio of the number of false alarms to the sum of false alarms and correct rejections; like DI, FA rate was computed separately at each stimulus position through the trial. The miss rate was calculated as Misses/(Hits+Misses). Variability in performance was calculated as the SD across sessions (Figs. 2B, 2C). Because misses occurred infrequently, miss rates for individual stimuli were computed within randomly selected blocks of 10 sessions, and the mean and SD were calculated across blocks (Fig. 4A).

2.5. Multidimensional scaling

The perceptual distance between stimuli was measured by constructing a matrix of 1-[FA rate] for each sample/nonmatch stimulus pair presented at position 2 (essentially the complement of those presented in Fig. 4B, below). Missing values from within-category comparisons that were not presented during the standard testing block were filled in by taking the average FA rate for a given stimulus across all sample and nonmatch presentations (i.e., averaging across the row and column that contained the missing value). The matrices were averaged across monkeys, and then averaged across the diagonal, to produce a single symmetrical distance matrix to which classic multidimensional scaling was applied (‘cmdscale’, MATLAB Statistics Toolbox).

2.6. Acoustic analysis

2.6.1. Rate and scale—The spectrotemporal characteristics of each stimulus were quantified using a two-stage computational model based on the neural tuning properties of the auditory periphery and cortex (Chi et al., 2005); MATLAB code obtained from http://www.isr.umd.edu/Labs/NSL/). This model has been applied previously to index the acoustic complexity of human and rhesus vocalizations (Joly et al., 2012), and our procedure largely follows theirs. Each sound was down-sampled to 16 kHz and converted to a spectrogram representation like those in Fig. 1 (parameters: 8 ms frame length, 8 ms time constant, linear function). The spectrogram corresponds to the frequency analysis performed by the peripheral auditory system, and serves as the input to the “cortical” stage: a bank of filters selective for the *scale* of spectral structure (the bandwidth of spectral modulation, from narrow to broad, in cycles/octave), as well as the *rate* of temporal modulation (the motion of spectral peaks, from slow to fast, in Hz). Rates spanned 6 to 40.4 Hz in quarter-octave steps (in both upward and downward directions), and scales spanned 0.25 to 13.45 cycles/octave in quarter-octave steps. The output of these filters is a representation in four dimensions: rate, scale, frequency, and time. Averaging across frequency, time, and scale yields a vector of filter outputs at each value of rate, i.e. a curve describing the power of temporal modulation in the stimulus at each rate. Upward and downward rates were folded together, and the centroid of this distribution was taken as a scalar measure of the temporal complexity of the sound, which we will refer to simply as “rate”. Likewise, averaging across frequency, time, and rate yields a distribution of scale; the centroid of this distribution was taken as a scalar measure of spectral complexity, and will be referred to as “scale”.

2.6.2. Spectral and temporal similarity—To determine if FA errors were related to the acoustic similarity between the nonmatch stimulus and the sample, the similarity of each sound pair was estimated in the spectral and temporal domains. All sounds were resampled to a common sampling rate of 32 kHz, and a spectrogram was generated using a 256-sample window with 50% overlap, at 129 linearly-spaced frequencies spanning 0 to 16 kHz (‘spectrogram’ function, MATLAB Signal Processing Toolbox; the same parameters were used to generate Fig. 1B). This function outputs a matrix ‘p’, the power spectral density across 129 frequencies, at each time point (the number of time points varied from 74 for 300-ms stimuli to 47 for the shortest sound). Summing across the rows of p yields an estimate of the frequency spectrum; summing down the columns of p generates a vector describing the total power over time, i.e. the envelope. Spectral similarity was measured as the Pearson correlation between the spectra of each sound pair. Temporal similarity between envelopes was calculated the same way, but, in cases where the sounds differed in length, the correlation was calculated by sliding the shorter envelope across the longer and measuring the correlation at each point; the maximum Pearson correlation was taken as the temporal similarity. Regression of spectral against temporal similarity for all nonidentical sound pairs confirmed that they were not collinear (p = 0.79), validating their use as independent predictor variables in the multiple linear regression described below.

Three additional measures of each sound's spectrum were extracted using Praat software (Boersma et al., 2012): (1) the centroid (on a log scale); (2) bandwidth (BW; standard deviation of the spectrum divided by the centroid); and (3) harmonic-to-noise ratio (HNR; mean periodicity by cross-correlation technique). The differences in these values for each stimulus pair were converted to degree of similarity (by subtraction from 1) and normalized to a range of 0 (identical) to −1 (most different of all pairs). Regression of each measure against the other two showed only a very weak relationship (R^2 values of 0.01, 0.07, and 0.02 for 1 vs. 2, 2 vs. 3, and 3 vs. 1, respectively). The latter two comparisons are significant at $p < 0.05$ (corrected for multiple comparisons), but on the basis of the weak R^2 values, all three measures were treated as independent variables in multiple linear regression.

3. Results

3.1 Serial DMS performance

Data were collected across 360 sessions for monkey F (>250,000 trials), and 116 sessions for monkey S (>82,000 trials). Both monkeys performed the serial DMS task at 67% correct overall and their performance varied across trial types (Figs. 2B, 2C). Relative to an earlier experiment that tested serial *visual* DMS with similar parameters ((Miller et al., 1993), performance on AA trials of auditory DMS was only slightly lower than performance on AA trials of visual DMS, but auditory DMS scores dropped off much more steeply than they did in vision as the number of nonmatch stimuli in the trial increased (cf. data from (Miller et al., 1993) overlaid on Fig. 2B). The effect was predominantly driven by an increase in FA rate between the second and third stimulus position, from ~0.15 to 0.5 in both subjects.

We reported earlier that accuracy on serial DMS was generally highest for trials in which a synthetic sound served as the sample (Scott et al., 2012). However, we hypothesized that the relevant stimulus attribute affecting memory performance was not strictly 'synthetic vs. natural', but rather that the synthetic sounds tended to be simpler than the natural sounds, both spectrally and temporally. We therefore analyzed the sounds in our stimulus set using a spectrotemporal method modeled on the tuning characteristics of auditory cortical neurons (Chi et al., 2005), in order to quantify the scales of spectral modulation and the rates of temporal modulation that were present in the sounds (see Methods, section 2.6.1). Among the sounds in our set, these rate and scale measures were strongly correlated ($r=0.91$, $p < 10^{-4}$), such that sounds tended to be either simple in both temporal and spectral domains (e.g., tones and noise), or complex in both (e.g., vocalizations, environmental sounds, and TORCs).

Plotting DI against the rate or scale of the sample stimulus confirms that performance was better for less complex sample stimuli in either domain (Fig. 3). The inverse relationship between performance and acoustic complexity was of at least borderline statistical significance in all cases (Monkey F, DI vs. rate: $R^2 = 0.25$, $p = 0.02$; DI vs. scale: $R^2 = 0.38$, $p = 0.003$; Monkey S, DI vs. rate: $R^2 = 0.18$, $p = 0.05$; DI vs. scale: $R^2 = 0.23$, $p = 0.03$, by linear regression).

3.2. Analysis of errors by stimulus

3.2.1. Miss errors—Although misses accounted for only a small proportion of total errors, misses occurred more often for certain stimuli (Fig. 4A). Thus, both animals made more miss errors for temporally complex stimuli (vocalizations and environmental sounds) than for simple stimuli (BPN, PT, and FM sweeps) The effect of sound category on miss rate was significant in both animals (one-way ANOVA, monkey F, F(6,245) = 110, $p < 10^{-4}$; monkey S, F(6,70) = 15.1, $p < 10^{-4}$). As was the case for overall performance, the only clear difference in miss rate between animals was to TORC stimuli, which were more frequently missed by monkey F than by monkey C.

3.2.2. False alarm errors—The majority of errors made by the monkeys were FA errors, in which the animal released the bar to a nonmatch sound. To ascertain whether FA errors correlated with particular stimuli, we computed the FA rates for each possible sample/nonmatch pair at the two positions where a nonmatch could occur (2 and 3). These FA rates are presented as a confusion matrix for each animal in Figures 4B and C. These matrices are strikingly similar for the two monkeys (monkey F, left column; monkey S, right column). Also notable is the symmetry along the diagonal, indicating that sample/nonmatch confusion was consistent irrespective of stimulus order. For each animal, there was a significant correlation between the patterns of FA rates at positions 2 and 3 (monkey F: r = 0.42, $p < 10^{-4}$; monkey S: r = 0.62, $p < 10^{-4}$). The patterns of errors at position 2 correlated significantly across animals (r = 0.53, $p < 10^{-4}$), but the noisier patterns at position 3 did not (r = 0.04, p = 0.78).

These error rates offer a window into the perceptual similarity among these stimuli, and the particular patterns of errors reveal some surprising effects. Both animals confused a band-passed noise (BPN) with a pure tone (PT) of the same center frequency (Fig. 4B, stimulus numbers for BPN: 4–6, PT: 7–9), and both confused environmental sounds with modulated noise (env: 19–21, TORC: 1–3). Perhaps more surprisingly, both monkeys confused conspecific vocalizations (Mvoc: 13–15) with sounds of several other categories, including TORCs and vocalizations of other species. To visualize perceptual similarity as inferred from the monkeys' FA errors, multidimensional scaling was applied to the confusion matrices at position 2 (Fig. 5). The dispersion of points along the first two MDS dimensions illustrates a clear divide between the spectrally and temporally simple stimuli (tones, noise, and FM) as opposed to complex stimuli, whether natural or synthetic.

3.3. Effects of intervening nonmatch stimuli

Confusion matrices were constructed to examine miss and FA rates at position 3 as a function of the similarity between the sample stimulus and the first (intervening) nonmatch stimulus at position 2 (Fig. 6A). Similarity between these first two stimuli may increase retroactive interference with memory of the sample and consequently increase the probability of an error in response to the subsequent stimulus at position 3.

3.3.1. Miss errors—Confusion matrices for FA responses (Figs. 4B, 4C) were symmetrical in appearance, suggesting that a given pair of stimuli were equally likely to be confused regardless of their order of presentation. Misses at position 3, by contrast, appeared to be more common when a temporally simple stimulus (e.g., PT or BPN) served as the intervening nonmatch stimulus, regardless of the sample (Fig. 6B). This effect did not hold if the order of stimuli was reversed (e.g., the miss rate for Mvocs was high following a PT distracter, but not vice-versa). The marginal distributions in Figure 6B clarify this disparity: Miss rate following a pure tone or bandpass noise distracter was above the mean (right side distributions), but miss rate for a PT or BPN sample was at or below the mean (upper distributions). In short, besides being better retained as sample stimuli, the simple sounds were also stronger nonmatch distracters.

3.3.2. False alarm errors—For both animals, FA errors at position 3 were distributed broadly across stimulus pairs, with fewer of the "hot spots" evident in position 2 (compare Figs. 4B and 6B). Errors tended to be lowest when both the sample and intervening nonmatch were of the simple, synthetic categories (PT, FM, or BPN), confirming that retention of these stimuli is relatively robust even after an intervening nonmatch.

3.4. False-alarms and acoustic similarity

The tendency to make FA errors during auditory DMS may be attributable to a level of acoustic similarity between the sample and nonmatch that is sufficient to exceed whatever internal threshold the listener may have set for a “match” response. This acoustic-similarity hypothesis leads to two predictions: first, a higher rate of FA errors would be expected for certain pairs of stimuli, as demonstrated above (Fig. 4B). The frequency distributions of FA errors illustrate the strong skew in the FA rate at position 2 (Fig. 7A, upper panel; same data as in Fig. 4B), such that most stimulus pairs had a very low FA rate, but a few pairs, which were frequently confused, formed a long tail on the right. The FA rate was about threefold higher at position 3 than at position 2 (Fig. 2C), and, whereas FA errors at position 2 were common after only a restricted subset of stimulus pairs, FA errors at position 3 were distributed more widely across stimuli (Fig. 7A, lower panel; same data as in Fig. 4C).

3.4.1. Multilinear regression—The second prediction of the acoustic similarity hypothesis is that FA errors will be more frequent for those pairs of sample/nonmatch stimuli that share the attribute by which subjects are determining a ‘match’ response. To test this, spectral and temporal similarity were measured as the correlation of the frequency spectra or temporal envelopes, respectively, of the two sounds in each pair (see Methods, section 2.6.2). The FA rate for each pair was subjected to multiple linear regression using spectral and temporal similarity as predictor variables. “Multiple” in the term above indicates that both similarity metrics served as inputs to the model and determined its predictive power (the R^2 value, or proportion of variance explained), and the relative contribution of each factor to the fit can be determined from its respective beta coefficient and p-value. At position 2 (Fig. 7B and upper portions of Fig. 7C), the only stimulus pair that may be tested is the sample and first nonmatch stimulus (S and NM1, respectively; data were log-transformed to approximate a normal distribution). At position 3, however, three comparisons can be made (lower portions of Fig. 7C): The similarity between: (i) the sample and NM1; (ii) the sample and NM2 (the sound that elicited the FA); and (iii) NM1 and NM2.

For FA errors at position 2, the regression accounted for a significant portion of the variance in FA rates of both animals, though the relationship was stronger in monkey F ($R^2 = 0.34$) than in monkey S ($R^2 = 0.12$). Comparisons of the beta coefficients and their respective p-values for the spectral and temporal regressors revealed that spectral similarity was the more powerful predictor of FA errors (the effect of temporal similarity was nonsignificant for monkey F and significant but relatively weak for monkey S; Table 1). To isolate which aspects of the spectrum may have been relevant to the monkeys’ matching behavior, the multilinear regression was run with the spectral centroid, BW, and HNR serving as the independent variables (Table 2). Both monkeys showed a strong effect of spectral centroid, but only monkey S showed an equally strong effect of BW; HNR was relatively weak for both.

Interpretation of FA errors at position 3 is more complex. For monkey F, FA rate was predicted equally well by the similarity of NM2 to either the original sample or the immediately preceding NM1 ($R^2 = 0.27$ and 0.26, respectively). The third comparison, between sample and NM1, explained about half as much of the variance ($R^2 = 0.13$) as had the two other similarity measures. By contrast, the FA rate for monkey S was predicted only by the similarity of NM2 to the immediately preceding NM1 ($R^2 = 0.12$, identical to variance explained at stimulus position 2). Similarity of NM2 or NM1 to the sample accounted for ≤ 3% of the variance in FA rate, at a borderline level of significance. As was the case at position 2, spectral similarity was the predominant factor in most comparisons. For monkey F, the similarity in spectral centroid between the nonmatch at position 2 and either of the preceding stimuli was the strongest predictor of FA rate, whereas monkey S appeared to employ similarity in BW between the nonmatch and sample as well.

The regression analysis was repeated using several indices of similarity derived from the rate and scale measurements (described in Methods, 2.6.1.). The effect of rate was nonsignificant in all cases, and the effect of scale was inconsistent, but in no case did the R^2 value exceed 0.09. In short, these higher-level metrics could not account for as much of the variance in FA rate as the simple correlation with frequency spectra.

3.4.2. Effect of stimulus class on similarity cues—Regression of FA rate against spectral and temporal similarity indicated that the former cue exerted a stronger influence on matching behavior. However, because three of our seven stimulus classes (PT, BPN, and FM) had essentially flat envelopes, our measure of temporal similarity may suffer from a ceiling effect, amplifying the apparent role of spectral similarity in matching behavior. To test this, we repeated the multilinear regression analysis at position 2 including only the subset of trials in which the sample and first non-match were both from the complex sound categories (TORC, voc, mvoc, and env). In monkey F, the results were substantially the same as those obtained from the full data set, in that the beta coefficient was significant only for spectral similarity (R^2 =0.11; $\beta_{spectral}$=0.55, p=0.002; $\beta_{temporal}$=n.s., p=0.72). In monkey S, spectral and temporal similarity both contributed to the explained variance in FA rate (R^2=0.17; $\beta_{spectral}$=1.05, p=0.001; $\beta_{temporal}$=1.24, $p<10^{-4}$).

3.4.3. Alternative test of the acoustic-similarity hypothesis—For an alternative test of the acoustic-similarity hypothesis, we set a threshold for FA rates significantly greater than chance and then compared the acoustic similarity of stimulus pairs that exceeded the threshold with those that did not. For each pair, a 95% confidence interval on the mean FA rate (across sessions) was estimated by bootstrap resampling (N = 1000 samples). If the lower bound of the confidence interval for a given pair was greater than the mean FA rate across all stimuli, the FA rate was considered to be significantly above chance. Acoustic similarity values for sound pairs with an FA rate significantly above chance were compared, as a population, against the similarity values of those pairs with FA rates indistinguishable from chance. For all comparisons in both animals, spectral similarity of the sound pairs that yielded above-chance FA rates was greater than the spectral similarity of the pairs that yielded chance FA rates (Wilcoxon rank-sum test, $p < 10^{-4}$ in all cases; same four comparisons as those used for the regressions in Table 1). The same comparisons for temporal similarity were significant in only one case (monkey S, at position 2, NM1 vs. sample, p = 0.006); all other p-values were >0.0125 (i.e., 0.05 after Bonferroni correction for four comparisons; Monkey F, NM1 vs. sample, p=0.55; NM2 vs. sample, p=0.88; NM2 vs. NM3, p=0.40; Monkey S, NM2 vs. sample, p=0.13; NM1 vs. NM2, p=0.04).

3.4.4. Behavioral validation of acoustic cues—The primacy of spectral over temporal cues in the monkeys' matching performance was tested in a control task, which was presented to monkey S only (monkey F was not available), by manipulating the stimuli to preserve either spectral features or temporal features, but not both (see Methods, section 2.3). DI at stimulus positions 2 and 3 dropped slightly with only spectral cues available, but dropped much more sharply with only temporal cues available, and in fact dropped to chance, i.e. below threshold, at position 3 (Fig. 8; comparisons of spectral only vs. temporal only, $p < 10^{-4}$ by Wilcoxon rank-sum test). The latter result suggests that the temporal features of the sensory trace may be more susceptible to interference than the spectral features and therefore less likely to be used as a delayed matching cue. However, we cannot rule out the possibility that the reliance on spectral cues was influenced by the available information in our stimulus set (see Discussion, section 4.2).

4. Discussion

We argued previously that the monkey's poor performance on auditory DMS reflects the absence of long-term memory in audition, and proposed that, in the auditory domain, nonhuman primates may be limited to sensory short-term memory (Scott et al., 2012). Specifically, we hypothesized that the tonotopic organization of the auditory system favors the retention of spectrally and temporally simple stimuli, and that matching errors resulted primarily from confusion due to spectral overlap between the sample and nonmatch stimulus. The present results support this interpretation, confirming that performance was better for spectrally and temporally simple stimuli, and that acoustic similarity, particularly in the spectral domain, can predict a significant portion of matching errors.

4.1. Effect of acoustic similarity

Auditory DMS performance has been reported to vary by sound type under some experimental conditions, but averaging acoustic measurements (modulation spectra and HNR) within categories failed to explain the performance difference between categories (Ng et al., 2009). We had moderate success explaining the variance in performance due to the sample sounds by using a representation that quantified both their spectral and temporal modulations (Fig. 3; (Chi et al., 2005; Joly et al., 2012). However, the more revealing approach was to apply acoustic similarity measures to the two-way analysis of FA errors by sample/nonmatch pair (Figs. 4 and 7).

Correct performance on our serial auditory DMS task required (i) retaining a short-term trace of the sample sound and (ii) comparing each subsequent test sound (match or nonmatch) to that sensory trace. The monkeys' imperfect performance (even in response to the very first nonmatch stimulus) implies that the short-term representation is imprecise and/or that the criterion applied by the animals to identify a match was a liberal one, thereby accentuating their bias towards FAs versus misses. The pattern of FA errors across stimulus pairs was similar for the two monkeys (Fig. 4), implying that they had similar response biases and employed similar match criteria. False alarm rate at stimulus position 2 was low for most sample/nonmatch pairs, with much higher FA rates for a small subset of pairs; but FA rate at the third position (after one intervening nonmatch) was much more widely distributed (Fig. 7A). This suggests that the nonmatch stimulus disrupted the short-term trace of the sample, effectively lowering the similarity criterion at which monkeys indicated a match.

We were able to account for a significant portion of the variance in FA rate across sample/nonmatch pairs using very simple metrics of acoustic spectral and temporal similarity. At stimulus position 2, where the nonmatch eliciting the FA followed immediately after the sample, acoustic similarity accounted for 34% and 12% of the variance in FA rate across stimulus pairs for monkeys F and S, respectively. At stimulus position 3, the rule by which animals were rewarded (viz., respond to the test stimulus identical to the sample) would predict that only similarity between the sample and second nonmatch would influence the probability of a FA. Alternatively, a pure "recency" effect would predict that similarity of the second nonmatch to the first nonmatch would also influence the probability of a FA. Monkey F seems to exhibit effects of both the rule and recency, inasmuch as this animal's response to the third stimulus is predicted equally well by similarity to either the sample or the first nonmatch. For monkey S, similarity between nonmatch 2 and nonmatch 1 accounts for 12% of the variance (the same result that was obtained for similarity between sample and nonmatch 1). However, similarity between the sample and nonmatch 2 accounts for much less of the variance; in effect, only similarity to the immediately preceding stimulus influences FA rate, which suggests a particularly strong effect of the intervening nonmatch in monkey S. For neither animal do the data support the possibility that they utilized only the sample's trace throughout the trial; rather, the data argue that their responses to a given test stimulus were based at least as much, if not more, on the immediately preceding stimulus within the trial.

Similarity between the sample (position 1) and first nonmatch (position 2) exerted only a weak effect on performance at position 3. For monkey F, this effect was about half as strong as the similarity of the stimulus at position 3 to either preceding stimulus, and, in monkey S, the influence was effectively absent. The effect in monkey F, though weak, suggests that an intervening nonmatch that was similar to the sample disrupted the trace of that sample more than did a nonmatch that was dissimilar to the sample. This effect has been demonstrated in human listeners for such acoustic attributes as pitch (Deutsch, 1972) and timbre (Starr et al., 1997) and may be due to overwriting by a stimulus that contains a feature similar to one contained in the sample (Lewandowsky et al., 2009).

One prior study of auditory memory in rhesus monkeys (Gaffan et al., 1991) used a confusion-matrix analysis similar to ours, with a similar outcome. In that study, the animals were trained to associate each of six different acoustic stimuli – two spoken words, and four simpler synthetic sounds – with one of six different visual stimuli. On each trial, 0.5 sec after the offset of an auditory stimulus, its correct visual associate was paired for choice with an incorrect associate. Though the task thus differed substantially from ours, the pattern of errors was consistent with the one we observed: performance was better for PT and FM stimuli overall, and the most common confusion was between the two words (i.e., the naturalistic stimuli). Apparently, the effect of acoustic complexity applies not only to purely auditory DMS but to cross-modal, auditory-to-visual DMS as well.

4.2. Acoustic features in pSTM

In most of the our analyses, only the spectral similarity between stimuli was a significant, albeit modest, factor in predicting error rate, the effect of temporal similarity being much weaker or nonsignificant (Section 3.4.1; Table 1). The control experiment supported the conclusion that the monkeys relied more on spectral cues, by demonstrating that their removal degraded behavior to a greater degree than removal of temporal cues (Section 3.4.4; Fig. 8). The spectral feature that best predicted matching errors was the centroid, which is effectively equivalent to the pitch of the remembered sound. (Pitch, as estimated by Praat software, correlated almost perfectly with the spectral centroid [$R^2 = 0.98$], but was undefined for 7 sounds that were broadband and noisy). The effect of bandwidth was also strong in monkey S, perhaps as a result of a difference in training histories between the animals: monkey F was naïve prior to training on DMS, but monkey S had been trained to discriminate a tone sequence from white noise and other broadband stimuli (Yin et al., 2008), a task for which bandwidth could have been exploited as a cue and associated with reward. However, both monkeys frequently confused PT and BPN stimuli of the same center frequency, despite their obvious qualitative differences to human listeners, implying that bandwidth was secondary to frequency or pitch in determining a match. The confusion of PT and broadband stimuli by rhesus monkeys in an operant task has been noted in another recent study (Kusmierek et al., 2012).

The relative weighting of pitch, bandwidth, and harmonicity in predicting performance varied not only between monkeys, but between stimulus positions 2 and 3. This is consistent with different features being maintained independently in pSTM, with the strongest effect being attributable to a pitch-specific memory mechanism unaffected by changes in other parameters (Demany et al., 2007). A similar mechanism has been proposed to exist in humans on the basis of behavioral studies suggesting that different auditory attributes are represented separately in auditory sensory memory, such as pitch and timbre (Semal and Demany, 1991) or pitch and loudness (Clement et al., 1999). The independence of features in auditory sensory memory is also supported by physiological studies of mismatch negativity (Nousak et al., 1996; Caclin et al., 2006). In this regard, the pSTM mechanism we propose differs from working memory and long-term memory, which rely on configural-or item-based, rather than feature-based, representations. The failure of the spectrotemporal rate and scale metrics to predict matching errors also aligns with the notion that pSTM exploits low-level features. For example, shifting the frequency of a sound by an octave does not change the sound's rate or scale, metrics that capture high-level features, such as harmonic structure and envelope modulation averaged across absolute frequency. The rate and scale metrics used here may be useful for exploring categorical representations (e.g., how a given word can be recognized regardless of the individual speaker), but these high-level metrics were inferior to simpler, low-level features in predicting pSTM performance (Results, 3.4).

The predominance of spectral similarity over temporal similarity in delayed-matching behavior may reflect a general property of auditory perception in monkeys, but we can not rule out the possibility that the design of our stimulus set influenced our subjects' strategy. When regression analysis was applied only to those trials in which both sample and nonmatch were complex (i.e., the subset of sounds with substantial variation in their temporal envelopes), monkey S showed an equal contribution of temporal and spectral similarity to matching behavior (see Results, 3.4.2). (Interestingly, the same monkey performed better than monkey F when the sample was a TORC stimulus, which are broad-band and must be distinguished by their temporal fluctuation.) This suggests that the monkey adapted his matching strategy to exploit the cues present in the stimuli, which for the full set would favor the spectral dimension. For example, after the stimuli were modified to remove either spectral or temporal information (see Sections 2.3, 3.4.4), 18 of 21 sounds retained distinct spectral profiles, but only 13 of 21 had distinct temporal envelopes. If a subject in the standard task were to attend to only one stimulus dimension, attending to the spectral dimension would allow better performance than would attending to the temporal dimension. In a similar vein, the short duration of the stimuli (~300 ms) limits the presence of slow modulations, possibly making temporal patterns the more difficult ones to extract. By contrast, spectral information can be extracted regardless of duration, and may be the easier dimension in which to retain and compare sounds (Scott et al., 2012). Future experiments would benefit from a careful balancing of putative acoustic features in the stimulus set, so as not to bias behavior in favor of one domain.

That monkeys may favor spectral over temporal cues in both auditory short-term memory and auditory discrimination was suggested by earlier studies. Thus, when *Cebus* monkeys were trained on an identity-matching paradigm at short delays, they learned to discriminate a high tone from a low tone, but failed to do so for a pulsed tone versus a static tone of the same frequency (Colombo et al., 1986; Lemus et al., 2009). It was later suggested that when challenged with discriminating complex auditory patterns that vary in multiple dimensions, monkeys may rely on local frequency differences rather than patterns (Colombo et al., 1996), p. 4513, citing (D'Amato et al., 1984; D'Amato et al., 1988). Even under more naturalistic conditions, frequency cues appear to dominate auditory perception: manipulation of Japanese macaque vocalizations revealed that amplitude modulation was of minor importance in discriminating two variants of their 'coo' call, which the monkeys could not consistently categorize after FM was removed (May et al., 1988). However, inasmuch as macaques have been successfully trained to exploit purely temporal cues in auditory delayed comparisons (Lemus et al., 2009), it is possible that although monkeys may more readily exploit frequency cues, they can adopt an alternative strategy if the stimulus set requires one. If, as we suggested above (Results, 3.4.2), the sensory trace of a rapidly fluctuating temporal signal is fragile and highly susceptible to retroactive interference, then reliance on frequency cues may be the best strategy that nonhuman primates and perhaps all vocal non-learners can use to compare auditory signals across a delay.

4.3 Conclusions

As noted earlier, recent studies have suggested that, in the auditory domain, monkeys may possess neither long-term memory (Fritz et al., 2005) nor even working memory (Scott et al., 2012), absence of the latter being consistent with the view that WM depends on the availability of representations stored in LTM. The residual, extremely impoverished form of auditory memory that remains, i.e., pSTM, stands in stark contrast to the monkey's visual memory, which consists of all three forms (pSTM, WM, and LTM). This marked mnemonic difference may be reflected in an anatomical difference between the monkey's auditory and visual systems. Encoding and storing the representations of stimulus items is known to depend on sensory input to the rhinal cortices (Murray et al., 2007), and whereas there is a dense projection to this division of the medial temporal lobe from the inferior temporal visual cortex, the projection from the superior temporal auditory cortex is relatively sparse (Munoz-Lopez et al., 2010; Suzuki et al., 1994).

Acknowledgments

We thank Helen Tak, Kathleen Moorhead, Peter Sergo, and Holly Vinal for assistance with animal training and data collection, and Alexander Kloth for programming assistance as well as data collection. This work was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services.

Abbreviations

LTM	long-term memory
STM	short-term memory
pSTM	passive short-term memory
WM	working memory
DMS	delayed-match-to-sample
DI	Discrimination index
FA	False Alarm

References

Boersma, P.; Weenink, D. Praat: doing phonetics by computer. 5.3.13. 2012.

Caclin A, Brattico E, Tervaniemi M, Naatanen R, Morlet D, Giard MH, McAdams S. Separate neural processing of timbre dimensions in auditory sensory memory. J Cogn Neurosci. 2006; 18:1959–72. [PubMed: 17129184]

Chi T, Ru P, Shamma SA. Multiresolution spectrotemporal analysis of complex sounds. J Acoust Soc Am. 2005; 118:887–906. [PubMed: 16158645]

Clement S, Demany L, Semal C. Memory for pitch versus memory for loudness. J Acoust Soc Am. 1999; 106:2805–11. [PubMed: 10573896]

Colombo M, D'Amato MR. A comparison of visual and auditory short-term memory in monkeys (Cebus apella). Q J Exp Psychol B. 1986; 38:425–48. [PubMed: 3809582]

Colombo M, Rodman HR, Gross CG. The effects of superior temporal cortex lesions on the processing and retention of auditory information in monkeys (Cebus apella). J Neurosci. 1996; 16:4501–17. [PubMed: 8699260]

D'Amato M, Colombo M. Auditory matching-to-sample in monkeys. Animal Learning & Behavior. 1985; 14:375–382.

D'Amato MR, Salmon DP. Processing and retention of complex auditory stimuli in monkeys (Cebus apella). Can J Psychol. 1984; 38:237–55. [PubMed: 6744117]

D'Amato MR, Colombo M. Representation of serial order in monkeys (Cebus apella). J Exp Psychol Anim Behav Process. 1988; 14:131–9. [PubMed: 3367099]

Demany, L.; Semal, C. The Role of Memory in Auditory Perception. In: Yost, WA.; Popper, AN.; Fay, RR., editors. Auditory Perception of Sound Sources. Vol. 29. Springer US; New York: 2007. p. 77-113.

Deutsch D. Mapping of interactions in the pitch memory store. Science. 1972; 175:1020–2. [PubMed: 5009395]

Fritz J, Mishkin M, Saunders RC. In search of an auditory engram. Proc Natl Acad Sci U S A. 2005; 102:9359–64. [PubMed: 15967995]

Gaffan D, Harrison S. Auditory-visual associations, hemispheric specialization and temporal-frontal interaction in the rhesus monkey. Brain. 1991; 114 (Pt 5):2133–44. [PubMed: 1933238]

Joly O, Ramus F, Pressnitzer D, Vanduffel W, Orban GA. Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. Cereb Cortex. 2012; 22:838–53. [PubMed: 21709178]

Kusmierek P, Ortiz M, Rauschecker JP. Sound-identity processing in early areas of the auditory ventral stream in the macaque. J Neurophysiol. 2012; 107:1123–41. [PubMed: 22131372]

Lemus L, Hernandez A, Romo R. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A. 2009; 106:9471–6. [PubMed: 19458263]

Lewandowsky S, Oberauer K, Brown GD. No temporal decay in verbal short-term memory. Trends Cogn Sci. 2009; 13:120–6. [PubMed: 19223224]

May B, Moody DB, Stebbins WC. The Significant Features of Japanese Macaque Coo Sounds - a Psychophysical Study. Anim Behav. 1988; 36:1432–1444.

Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci. 1993; 13:1460–78. [PubMed: 8463829]

Mishkin M. Memory in monkeys severely impaired by combined but not by separate removal of amygdala and hippocampus. Nature. 1978; 273:297–8. [PubMed: 418358]

Munoz-Lopez M, Mohedano-Moriano A, Insausti R. Anatomical Pathways for Auditory Memory in Primates. Frontiers in Neuroanatomy. 2010; 4:129. [PubMed: 20976037]

Murray EA, Mishkin M. Severe tactual memory deficits in monkeys after combined removal of the amygdala and hippocampus. Brain Research. 1983; 270:340–344. [PubMed: 6883103]

Murray EA, Bussey TJ, Saksida LM. Visual perception and memory: a new view of medial temporal lobe function in primates and rodents. Annu Rev Neurosci. 2007; 30:99–122. [PubMed: 17417938]

Ng CW, Plakke B, Poremba A. Primate auditory recognition memory performance varies with sound type. Hear Res. 2009; 256:64–74. [PubMed: 19567264]

Nousak JM, Deacon D, Ritter W, Vaughan HG Jr. Storage of information in transient auditory memory. Brain Res Cogn Brain Res. 1996; 4:305–17. [PubMed: 8957572]

Scott BH, Mishkin M, Yin P. Monkeys have a limited form of short-term memory in audition. Proc Natl Acad Sci U S A. 2012; 109:12237–41. [PubMed: 22778411]

Semal C, Demany L. Dissociation of Pitch from Timbre in Auditory Short-Term-Memory. Journal of the Acoustical Society of America. 1991; 89:2404–2410. [PubMed: 1861000]

Starr GE, Pitt MA. Interference effects in short-term memory for timbre. J Acoust Soc Am. 1997; 102:486–94. [PubMed: 9228812]

Suzuki WL, Amaral DG. Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. The Journal of Comparative Neurology. 1994; 350:497–533. [PubMed: 7890828]

Wright AA. Auditory list memory and interference processes in monkeys. J Exp Psychol Anim Behav Process. 1999; 25:284–96. [PubMed: 10423854]

Yin P, Fritz JB, Shamma SA. Do ferrets perceive relative pitch? J Acoust Soc Am. 2010; 127:1673–80. [PubMed: 20329865]

Yin P, Mishkin M, Sutter M, Fritz JB. Early stages of melody processing: stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R. J Neurophysiol. 2008; 100:3009–29. [PubMed: 18842950]

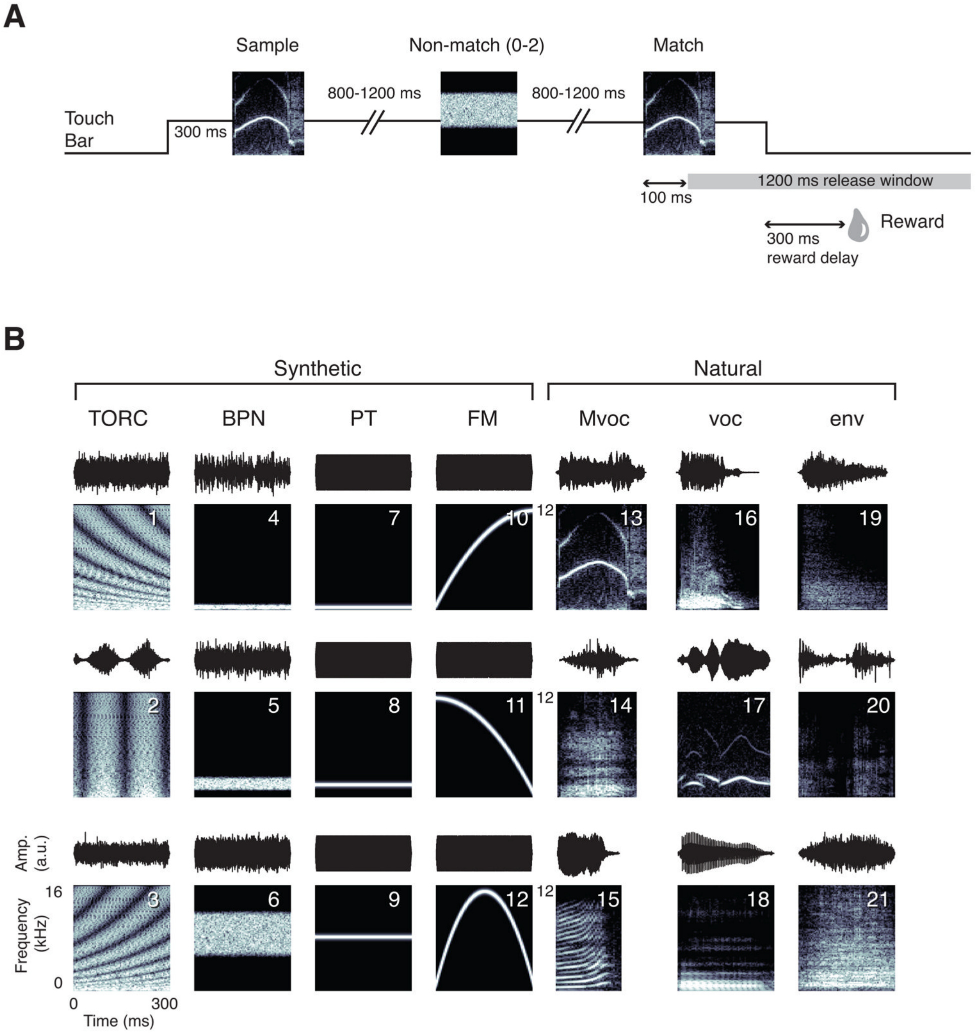

Figure 1.

(A) Schematic diagram of the timing of a delayed match-to-sample (DMS) trial (see methods). **(B)** Set of 21 sounds used during DMS testing. The sounds are illustrated as both amplitude-time waveforms (upper panel of each panel pair) and as frequency-time spectrograms (lower panel of each pair). The set includes three exemplars (rows) for each of seven categories (columns): (1) temporally orthogonal ripple complexes (TORCs); (2) 1/3-octave band-pass noise (BPN) at center frequencies of 512, 2048, and 8192 Hz; (3) pure tones (PT) at the same frequencies; (4) frequency-modulated sweeps (FM), upward, downward, and bi-directional, between 0.25 and 16 kHz; (5) rhesus monkey vocalizations (Mvoc), archscream, bark, and coo; (6) other species' vocalizations (voc), dog bark, bird song, and vowel voiced by human female /a/; (7) environmental sounds (env), cage door closing, click of water solenoid opening, and metallic noise. All synthetic sounds were 300 ms in duration, whereas the natural sounds varied in duration, with the Mvocs tending to be shorter than the other categories (stimulus 13, 282 ms; 14, 246 ms; 15, 195 ms; 16, 257 ms; 17, 288 ms; 18, 300 ms; 19, 280 ms; 20 and 21, 300 ms each. The frequency axis on all spectrograms spans 0–16 kHz (linear scale), except those of the three Mvocs, for which the axis spans 0–12 kHz.

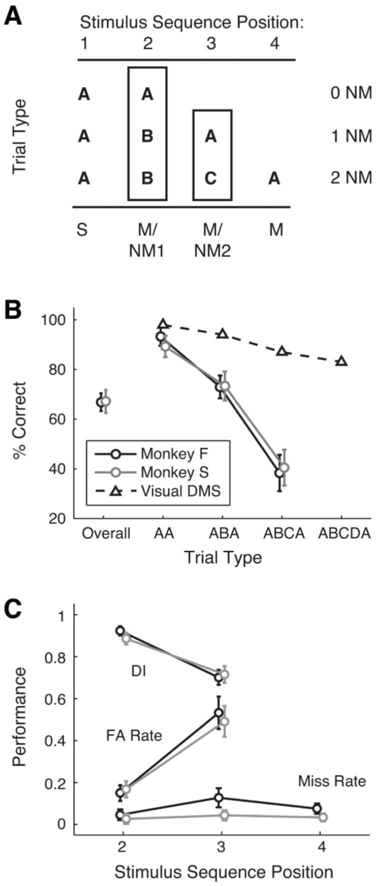

Figure 2.

(A) Schematic diagram of stimulus sequence positions (numbered across the top), and the stimulus conditions that may appear within the three trial types (AA, ABA, and ABCA, corresponding to zero, one, or two nonmatch stimuli). The stimulus at position 1 is always the sample (S); stimuli at positions 2 and 3 may be a match (M) or a nonmatch (NM1, NM2); position 4 is always a match. **(B)** Percent correct (mean ±SD) for all trials (Overall, at left) and for the three trial types (AA, ABA, and ABCA, at right). Black symbols represent monkey F (N = 360 sessions, > 250,000 trials); gray symbols represent monkey S (N = 116 sessions, > 82,000 trials). For comparison, performance on an analogous visual DMS task is overlaid (open triangles and dashed line; from Miller et al., 1993). **(C)** Performance measured by DI, FA rate, and miss rate, computed separately at each position within the trial sequence. All three metrics take a value between 0 and 1 and share the same ordinate. FA rate and DI are not computed for the stimulus at position 4, as it is always a match, and so no FA can occur. FA rate increases sharply between stimulus positions 2 and 3 for both monkey F (black) and monkey S (gray).

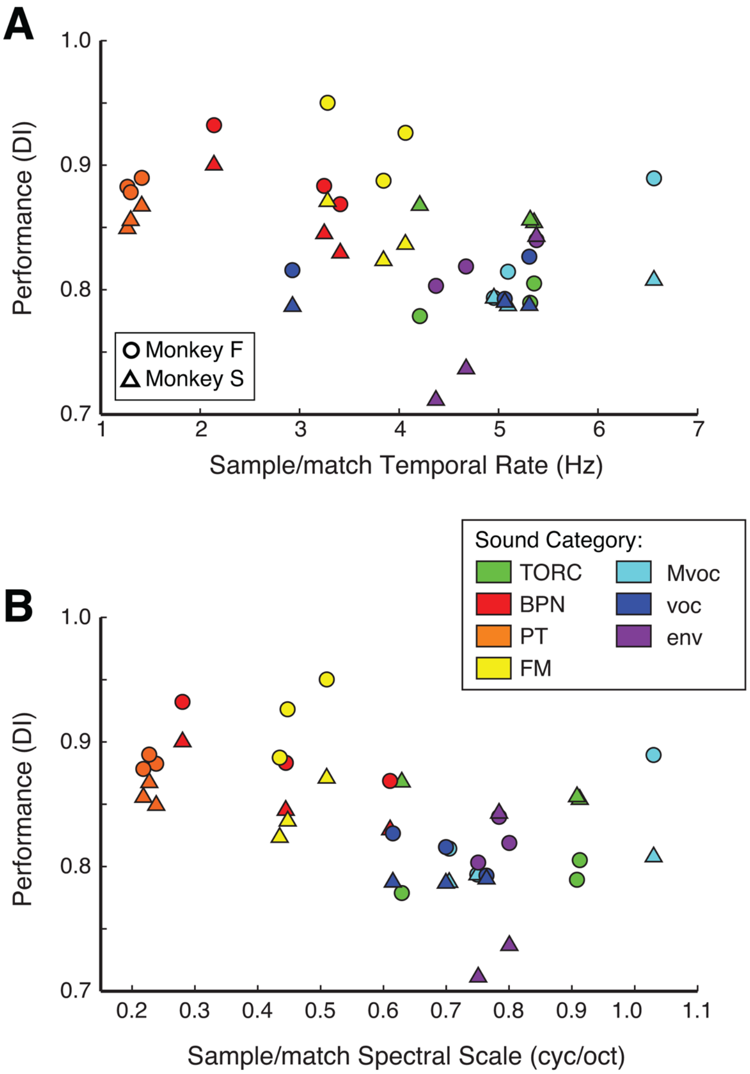

Figure 3.

Performance (DI) compared directly to temporal rate (A) and spectral scale (B) confirms that performance decreases as sample sound complexity increases. Circles represent data from monkey F, triangles, from monkey S. Colors indicate the experimenter-defined “categories”, where hot colors indicate simple synthetic stimuli, and cool colors indicate complex natural stimuli. The negative correlation of performance with complexity is significant (p ≤ 0.05) by linear regression in all cases.

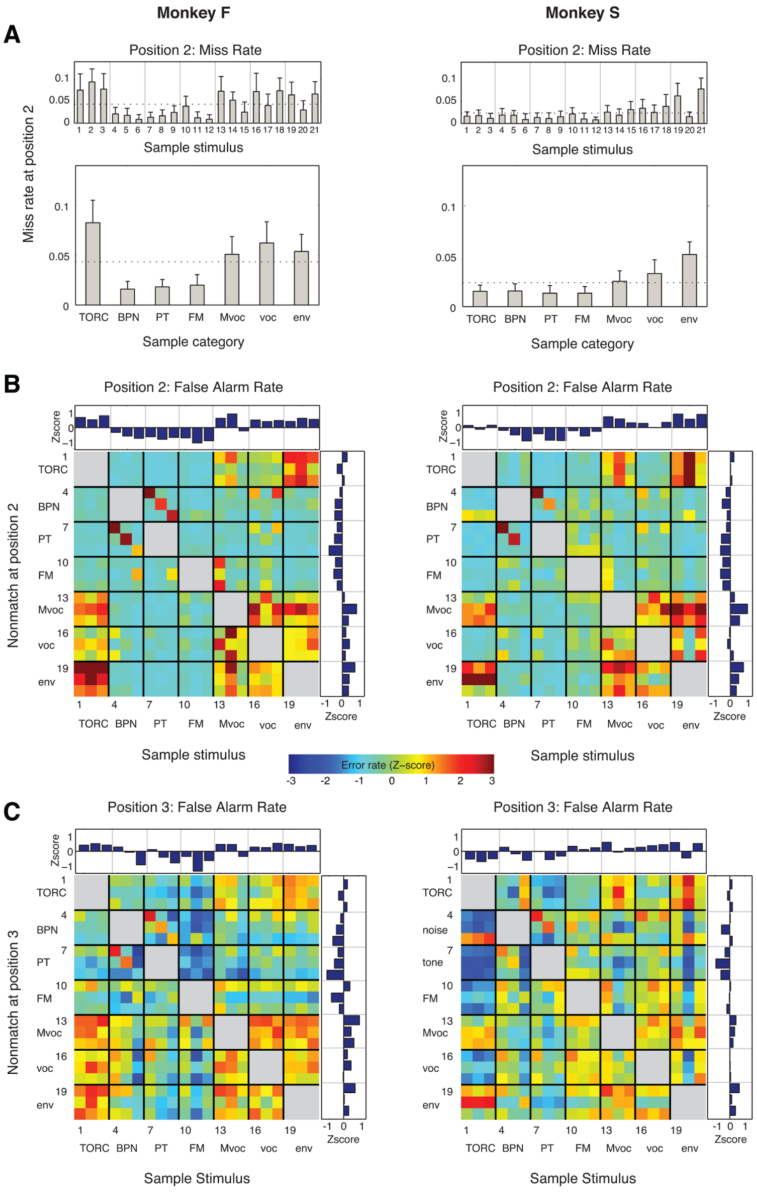

Figure 4.

Miss and FA rates as a function of stimulus type. (A) Miss error rate (mean +SD) at position 2 for individual stimuli (upper panels) and averaged by sound category (lower panels); dotted line represents the mean across all stimuli, and vertical gray lines separate sound categories in the upper panel. In this and subsequent panels, data on the left and right are from monkeys F and S, respectively (see also Fig. 1 for identity and category of the numbered sounds). (B) Confusion matrices showing the FA rate for each possible pair of sample (columns) and immediately following nonmatch (rows) at stimulus sequence position 2. FA rate on the color axis is plotted as a Z-score normalized to SDs relative to the mean FA rate across all stimuli at this position. Sound categories are labeled on the margins, and numerals refer to the stimulus numbers in Figure 1. Sample and nonmatch stimuli were never from the same category, indicated by the gray blanks along the diagonal. Insets above and to the right of the matrix are marginal distributions (e.g. the upper inset plots FA rate by sample, averaged across all nonmatch stimuli). The mean number of trials per stimulus pair is 427 for monkey F (left panel), and 147 for monkey S (right panel). **(C)** Same conventions as those for panel B, but for FA errors at stimulus position 3 (i.e., after the animal has correctly withheld responding to the first nonmatch stimulus, at position 2). Mean trials per stimulus pair: 181 for monkey F, 62 for monkey S.

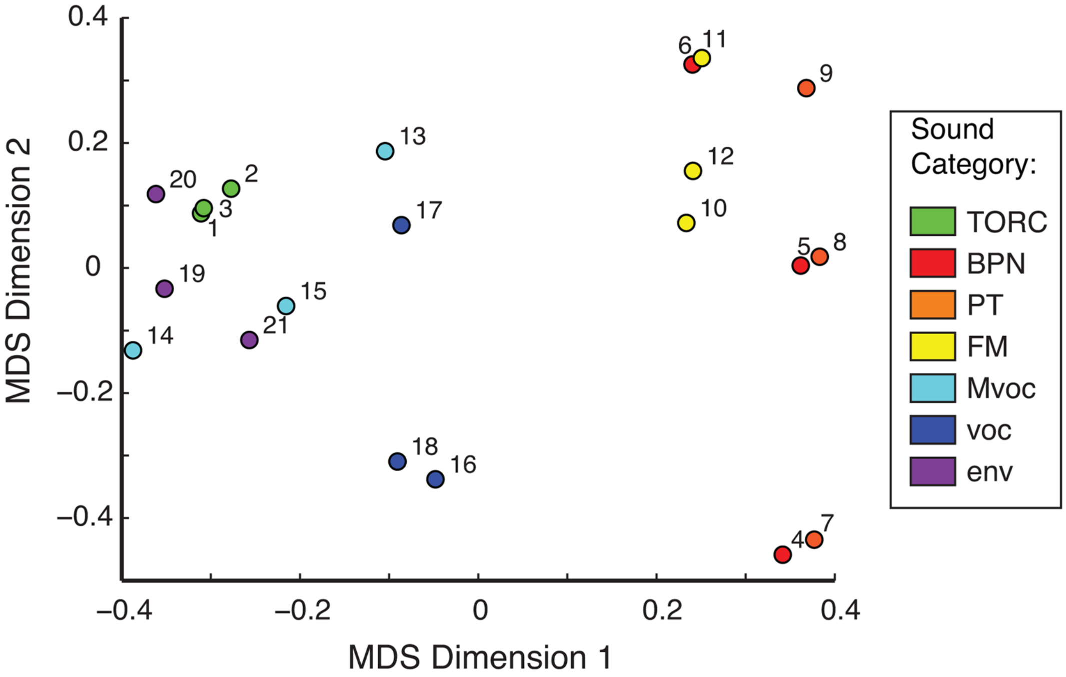

Figure 5.

Multidimensional scaling of stimulus distance based on FA rate. Note wide gap between simple and complex stimuli (hot and cool colors, respectively) and the strong perceived similarity between tones and band-pass noise of similar frequency (e.g., 4 and 7, 5 and 8).

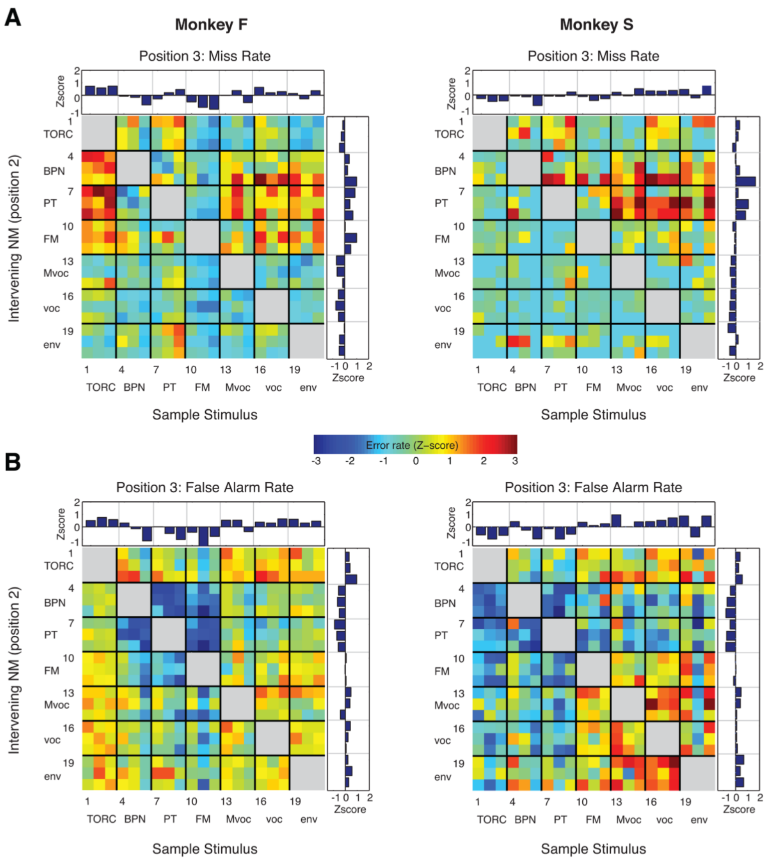

Figure 6.

Effect of the intervening nonmatch stimulus on miss and FA rates. (A) Distribution of miss errors at stimulus position 3 as a function of both the sample and the first nonmatch. Whereas confusion matrices for FA errors were symmetric in appearance (i.e., the order of presentation was irrelevant), miss errors were more common after presentation of a nonmatch consisting of a PT or BPN, suggesting that these were potent distracters regardless of the sample stimulus. Conventions the same as those in Figure 4A. (B) Distribution of FA errors at stimulus position 3 as a function of both the sample and the first nonmatch (the intervening stimulus at position 2, to which the animal correctly withheld responding). Mean trials per stimulus pair: 181 for monkey F, 62 for monkey S. Conventions the same as those in Figure 4B.

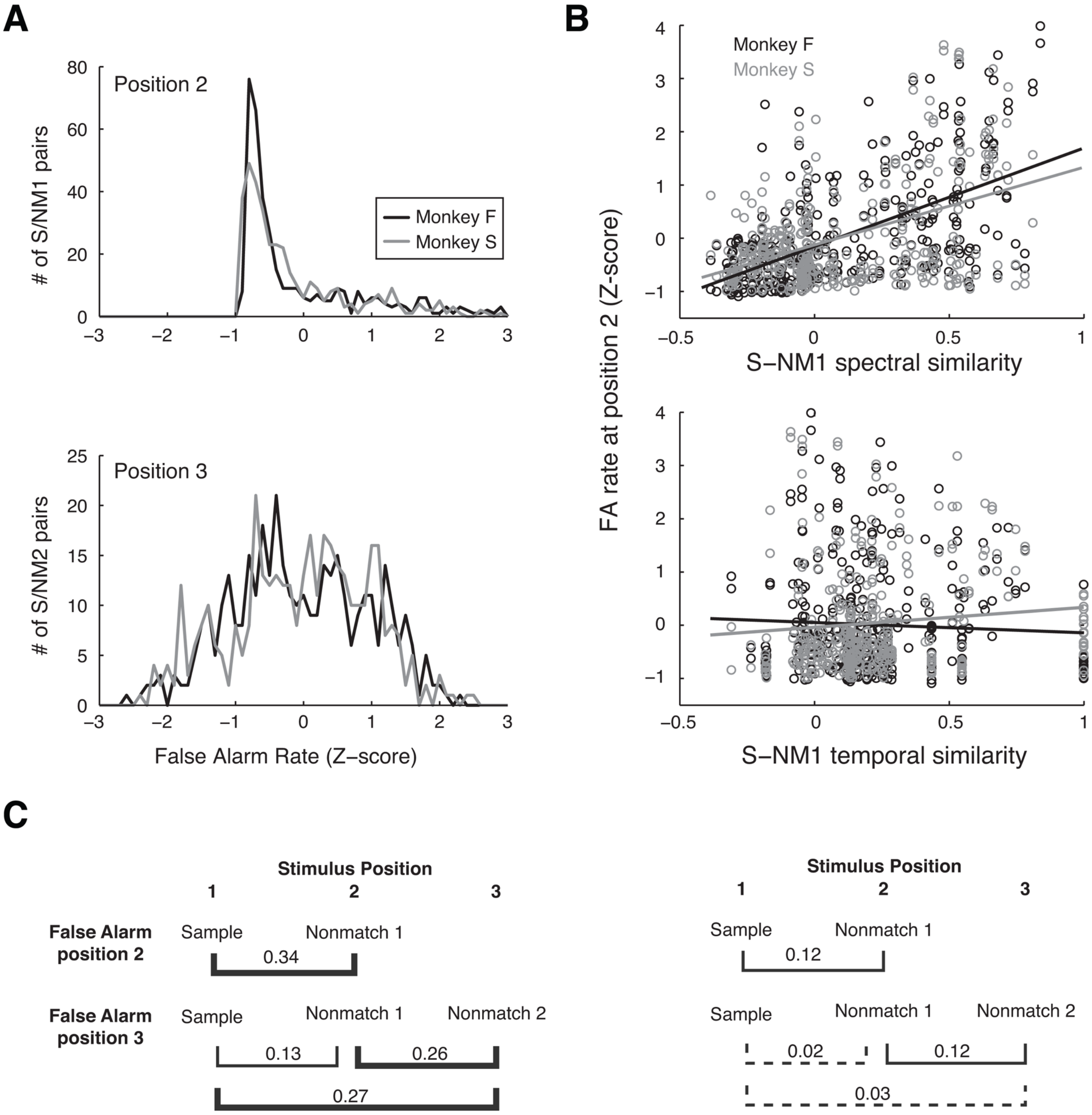

Figure 7.

(A) Distribution of FA rate across all possible sample/nonmatch pairs at stimulus positions 2 and 3 (upper and lower panels, respectively). FA rate is plotted as a Z-score normalized to SDs relative to the mean FA rate across all stimuli at that position for monkeys F and S (black and gray curves, respectively). The skewed shape and long tail at stimulus position 2 reflects the finding that a few sound pairs elicited FAs frequently, but most sound pairs did not (same data as Fig. 4B; see section 3.2.2 for discussion of frequently-confused sound pairs). (B) Scatter plots of FA rate at stimulus position 2 as a function of the spectral similarity (upper panel) and temporal similarity (lower panel) between all possible sample/nonmatch pairs. Slopes of fits from linear regression for spectral similarity are 1.9 and 1.5 for monkey F (in black) and S (in gray), respectively; and for temporal similarity they are −0.2 and 0.3 for monkeys F and S, respectively. (For illustration, regression was performed individually for each predictor; these slopes are slightly different from the beta values in Table 1, which were derived from multiple linear regression using both predictors; see Methods for computation of spectral and temporal similarity.) (C) R^2 values from multiple linear regression of FA rate and stimulus similarity for monkeys F and S (left and right panels, respectively). Each bracket links the pair of stimulus positions for which FA rate was regressed against both spectral and temporal similarity; line weight represents the percent of the variance accounted for by the regression model (brackets in bold, > 25%, significant; thin brackets, < 25%, significant; dashed brackets, nonsignificant).

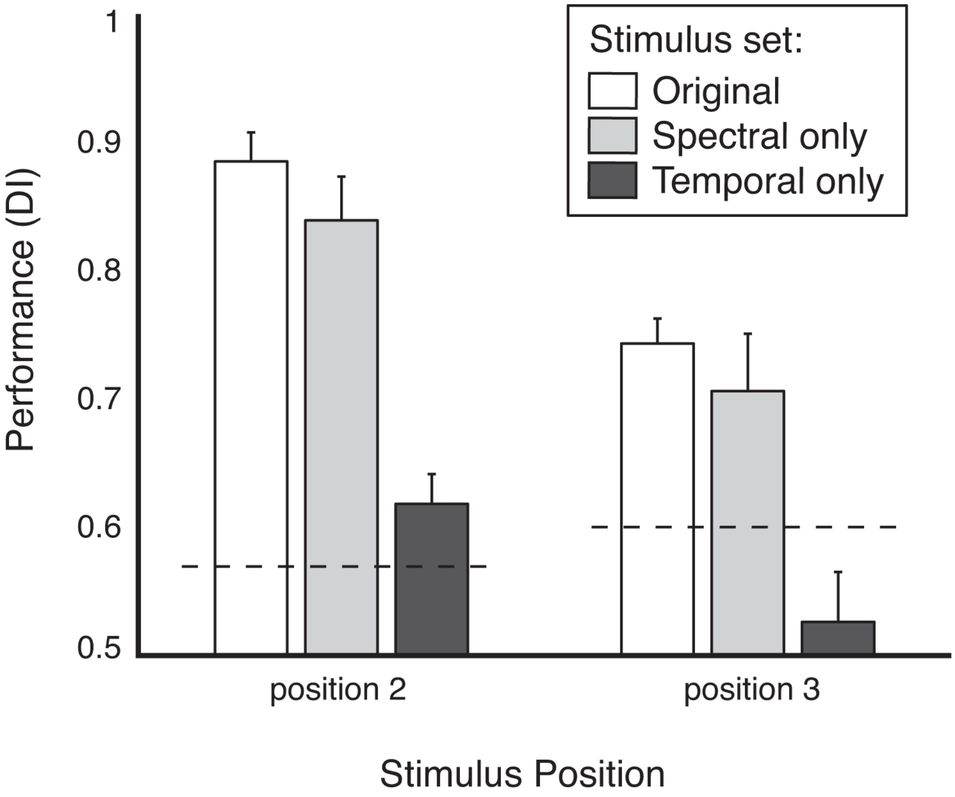

Figure 8.

Task performance using a modified set of stimuli that had either preserved spectral cues but identical temporal envelopes (light gray bars) or preserved temporal envelopes but identical spectra (dark gray bars). Data are from monkey S (monkey F was not available for testing). The performance data for spectral cues only were gathered in 18 sessions (10,416 trials), for temporal cues only, in 16 sessions (12,989 trials), and for both cues (i.e., unmodified stimuli) in 16 sessions (10,380 trials), the latter sessions selected to include those closest in time to the sessions with the modified stimuli. Data are shown as DI (mean +SD) calculated separately for stimulus positions 2 and 3. Dashed lines indicate threshold performance. All differences within the same position are significant at $p < 10^{-4}$, except for the difference between original and spectral cues only at position 3, which is significant at $p = 0.02$ (Wilcoxon rank-sum test). Performance with temporal cues only at stimulus position 3 did not exceed chance.

Table 1

Multiple linear regression of FA rate by acoustic similarity

	Position 2:				
	NM1 vs. Sample				
Subject	**R^2**	**$\beta_{spectral}$**	**$p_{spectral}$**	**$\beta_{temporal}$**	**$p_{temporal}$**
F	0.34	1.4	$< 10^{-4}$	n.s.	0.71
S	0.12	1.1	$< 10^{-4}$	0.6	0.002
	Position 3:				
	NM1 vs. Sample				
	R^2	$\beta_{spectral}$	$p_{spectral}$	$\beta_{temporal}$	$p_{temporal}$
F	0.13	0.09	$< 10^{-4}$	−0.09	$< 10^{-4}$
S	0.02	0.06	0.008	n.s.	0.77
	NM2 vs. Sample				
	R^2	$\beta_{spectral}$	$p_{spectral}$	$\beta_{temporal}$	$p_{temporal}$
F	0.27	0.21	$< 10^{-4}$	n.s.	0.92
S	0.03	0.08	0.01	0.09	0.01
	NM1 vs. NM2				
	R^2	$\beta_{spectral}$	$p_{spectral}$	$\beta_{temporal}$	$p_{temporal}$
F	0.26	0.21	$< 10^{-4}$	n.s.	0.07
S	0.12	0.15	$< 10^{-4}$	n.s.	0.77

Table 2

Multiple linear regression of FA rate by spectral similarity

	Position 2:						
	NM1 vs Sample						
Subject	**R^2**	**$\beta_{centroid}$**	**$p_{centroid}$**	**β_{BW}**	**p_{BW}**	**β_{harm}**	**p_{harm}**
F	0.31	1.47	$< 10^{-4}$	0.41	0.002	0.56	$< 10^{-4}$
S	0.37	1.46	$< 10^{-4}$	1.49	$< 10^{-4}$	0.87	$< 10^{-4}$
	Position 3:						
	NM1 vs. Sample						
	R^2	$\beta_{centroid}$	$p_{centroid}$	β_{BW}	p_{BW}	β_{harm}	p_{harm}
F	0.21	n.s.	0.88	−0.13	$< 10^{-4}$	0.15	$< 10^{-4}$
S	0.12	n.s.	0.97	0.07	0.009	0.13	$< 10^{-4}$
	NM2 vs. Sample						
	R^2	$\beta_{centroid}$	$p_{centroid}$	β_{BW}	p_{BW}	β_{harm}	p_{harm}
F	0.28	0.22	$< 10^{-4}$	0.05	0.04	0.11	$< 10^{-4}$
S	0.32	0.10	0.006	0.33	$< 10^{-4}$	0.16	$< 10^{-4}$
	NM1 vs. NM2						
	R^2	$\beta_{centroid}$	$p_{centroid}$	β_{BW}	p_{BW}	β_{harm}	p_{harm}
F	0.47	0.30	$< 10^{-4}$	0.17	$< 10^{-4}$	0.09	$< 10^{-4}$
S	0.23	0.15	$< 10^{-4}$	n.s.	0.05	0.16	$< 10^{-4}$